

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: October 1, 1999, 15:32:18 ; Search time 1619.02 Seconds
(without alignments)
3751.894 Million cell updates/sec

Title: US-09-026-400-3
Perfect score: 1910
Sequence: 1 cgcgctactagtagtattcc.....aagtaaaaaaaaaaaaaa 1910

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_em:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_on:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: gb_ba1:*
39: gb_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	144.8	7.6	93695	7	ATF2009		AL021749 Arabidops

2	142.2	7.4	103495	8	ATAC006585		AC006585 Arabidops
3	140.6	7.4	2362	12	RNTATR		X02741 Rat mRNA fo
4	134.2	7.0	61384	8	ATAC007048		AC007048 Arabidops
5	131	6.9	2004	12	RATATTRA		M18340 Rat tyrosin
6	121	6.3	39524	1	SCH10		AL049754 Streptomy
7	118	6.2	2754	10	HSTATR		X52320 Human mRNA
8	118	6.2	2051	10	HSTYAT		X55675 H.sapiens m
9	116.8	6.1	73840	7	AB007644		AB007644 Arabidops
10	112.4	5.9	25970	1	SC2H4		AL031514 Streptomy
11	111	5.8	39399	1	MSGB1970CS		L78815 Mycobacteri
12	110.2	5.8	70475	7	AB018112		AB018112 Arabidops
13	103.6	5.4	7409	1	STMVBRAL		D50624 Streptomyce
14	102.6	5.4	9150	1	MTVCY279		Z97991 Mycobacteri
15	102.2	5.4	3240	1	SAAJ5198		AB005198 Streptomy
16	102.2	5.4	3240	1	SAN6517		AB006517 Streptomy
17	102.2	5.4	119430	7	ATF9D16		AL035394 Arabidops
18	97.2	5.1	47852	1	MTV023		AL022022 Mycobacte
19	93.8	4.9	29550	1	MTV039		AL021942 Mycobacte
20	92.6	4.8	53662	1	MTV016		AL021841 Mycobacte
21	88.6	4.6	2767	7	SSI224970		AJ224970 Spermatoz
22	88.4	4.6	32258	36	CEFA2D1		Z81081 Caenorhabdi
23	87.6	4.6	187647	34	CEX16B4		AL021148 Caenorhab
24	86.4	4.5	67200	1	MTV017		AL021897 Mycobacte
25	85	4.5	2764	17	HSBBI4P4B		L14321 Bovine herp
26	84.2	4.4	40790	1	MTCY493		Z95844 Mycobacteri
27	83.4	4.4	3619	1	SCVALSFP		Y13070 S.coelicolo
28	80.4	4.2	1777	1	TTHASPC		D38459 Thermus the
29	80.2	4.2	67200	1	MTV017		AL021897 Mycobacte
30	80.2	4.2	135301	17	BHV1CGEN		AJ004801 Bovine he
31	80.2	4.2	135301	17	BHV1CGEN		AJ004801 Bovine he
32	80.2	4.2	8113	17	HSBBI4P4A		L14320 Bovine herp
33	79.2	4.1	7585	1	SCY14206		Y14206 Streptomyce
34	78.8	4.1	28826	1	MTV041		AL021958 Mycobacte
35	78.6	4.1	131993	17	OP075930		U75930 Orgyia pseu
36	78	4.1	6943	2	AF072135		AF072135 Caulobact
37	77.8	4.1	26440	1	SCE15		AL049707 Streptomy
38	77.6	4.1	47852	1	MTV023		AL022022 Mycobacte
39	77.2	4.0	749	17	NPORPT2		D13755 Multicapsid
40	76.2	4.0	31300	1	MTV035		AL021930 Mycobacte
41	76	4.0	53662	1	MTV016		AL021841 Mycobacte
42	75.8	4.0	102353	35	AC007708		AC007708 Homo sapi
43	75.4	3.9	16150	1	MTV044		AL021999 Mycobacte
44	75.4	3.9	32039	1	SC7A1		AL034447 Streptomy
45	75.4	3.9	3808	2	AF068743		AF068743 Haloferax

ALIGNMENTS

RESULT	1	ATF2009	93695 bp	DNA	PLN	27-OCT-1998
LOCUS		Arabidopsis thaliana				
DEFINITION		Arabidopsis thaliana DNA chromosome 4, BAC clone F2009 (ESSAII project).				
ACCESSION		AL021749				
NID		g2842474				
VERSION		AL021749.1				
KEYWORDS		GI:2842474				
SOURCE		thale cress.				
ORGANISM		Arabidopsis thaliana				

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
1 (bases 1 to 93695)
Bevan,M., Koetter,P., Hempel,S., Entian,K.-D., Hoheisel,J.,
Mewes,H.W., Mayer,K.F.X. and Schueller,C.
Unpublished

2 (bases 1 to 93695)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (16-OCT-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge

Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk

FEATURES

source

Location/Qualifiers

1..93695

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

18..467

/gene="F2009.10"

complement(18..467)

/gene="F2009.10"

/number=1

complement(18..467)

/gene="F2009.10"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAAL6872.1"

/db_xref="PID:e1250050"

/db_xref="PID:g2842475"

/db_xref="GI:2842475"

/db_xref="SPTREMBL:O49442"

/translation="MTSKFLRIVSKRRTWACLFVYVAILLSSSNLSINWYIG
ENHOTSSGLPAAIVASVLGAVFGVLSMAALFIAVPAIVVWIVVYVIMAFAGSKRR
VWIEGRKVTKEIAGFVFRVLKEGNFALLCALLAYFFFNYSYSSS"

complement(2196..2578)

/gene="F2009.20"

/number=1

2196..4222

/gene="F2009.20"

complement(join(2196..2578,3325..3476,4053..4222))

/gene="F2009.20"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAAL6873.1"

/db_xref="PID:e1250051"

/db_xref="PID:g2842476"

/db_xref="GI:2842476"

/db_xref="SPTREMBL:O49443"

/translation="MNSMPVVDNRKSRDAREKRMTLENMTGERGFVDVCRCEVLW
EEEEELCKRDCASITNQSLITLVEYVRLGDDVDSVOHKCYCTIHNFKPTDECY
LPFLIVALLSSRNLSINWYIGENHTSSGLPAAIVASVLGAVFGVLSMAALF
IAVPAIVVWIVVYVIMAFAGSKRRVWIEGRKVTKEIAGFVFRVLKEGNFALLCALLAYFFFNYSYSSS"

complement(2579..3324)

/gene="F2009.20"

/number=1

complement(3325..3476)

/gene="F2009.20"

/number=2

complement(3477..4052)

/gene="F2009.20"

/number=2

complement(4053..4222)

/gene="F2009.20"

/number=3

4143..4941

/gene="F2009.40"

/number=1

4143..6188

/gene="F2009.40"

join(4143..4941,4987..5044,5096..6188)

/gene="F2009.40"

/note="similarity to lectin receptor-like serine/threonine
kinase lecrK1, Arabidopsis thaliana, PIR2:S68589; Contains
Protein kinases signatures and profile; pos. 328-349 &
pos. 444-456"

/codon_start=1

/product="receptor protein kinase like protein"

/protein_id="CAAL6875.1"

/db_xref="PID:e1250053"

/db_xref="PID:g2842478"

/db_xref="GI:2842478"

/db_xref="SPTREMBL:O49445"

/translation="MFSKVILLFSLASLLLRSTTGTGIEFIYNSFTTTTLLGNAT
VKSPPSILTTNQTTSIGRGLYPSRINASSSSASPLPEATSRIFSMAFKHLSPGHG
FAVFLPFTSETASASSOHLGLNFNTNNGDPNSRIFAEDVFANQFNINDNHVG
DVNSLTVSATETAGFYGRDQGRFTLKLNGENYQAWIEFNQSAINTWARASSRKP
IRPLISPLNLTGLVDDMEFTASTGOLVQSHRILSWFSNSNFSIGDSVLKSKGF
IAGVSSGWLQRLGDEWETEPHRYQKDVLEATKGFSDENMIGYGGNSKYRG
VLEGEVAVKRIMSPRESVGTSEFLAEVSSGLRLRHKNIIVGLKWSKGGESLILI
KEYMNGSDKRLFDCCNMLNWEERMEVIRDLASGLMLYHGHETKVLHQRDKSSNVL
LDKDMNARVGDGLAKLQNTSKENWSTTHVVTAGYMAPELVKTGRASQTDVYSGV
FVLEVCGRRIPIEGREGIVEWLMEKDKVVDGLDERIKANGVFVVEVEWALRIG
LLCVHPDPVRYRPKRQVQILEOGRVLVDGGEREISLLERVKSYLLETGEGSRQHP
TFQDVWNSYSNSFQTYDSILHGR"

complement(4494..4566)

/gene="F2009.40"

/number=1

4494..11254

/gene="F2009.30"

complement(join(4494..4665,6567..6669,6878..7149,
7305..7371,7537..7647,7763..7915,9901..9975,10605..10996,
11074..11254))

/gene="F2009.30"

/note="similarity to ENOD20 gene (small copper-binding
protein), Medicago truncatula, PATCHX:E258656"

/codon_start=1

/product="copper-binding protein-like"

/protein_id="CAAL6874.1"

/db_xref="PID:e1250052"

/db_xref="PID:g2842477"

/db_xref="GI:2842477"

/db_xref="SPTREMBL:O49444"

/translation="MGLVMRFDLVLMFVLMGLGFTISNGYKFYVGGKDWVPTPSED
YSHSHENRFQVNDTLHFYAKGKDSVLETEQYNTCTHPLTSLSDGDSLPLLSH
SGSYFFISGNSQCLKGOKLAVKLVSTVHSHSPRHTSPSPVHQHLSLSPGSPGVE
PSSDSNRVPAPGPATAPNSAGLVPGMVVIVVNSIIPKIDSLKLSLPLVDVAVG1
STSRQQAEEPPVSSPLSGLSGKEEQKIIPKQKVAVLKS1KQSPKKNVLA
ALVGRMVEDALIQLVTVRAAQTVYRVIIHAARANATHNGLDPPDLVAEAFVKG
LFGKVVYHAKGRGIIISIPCRLTIVIVRETTPEEAEIARLKHVNEKKSKRRLV
PKLIETSPWNRGRTKANRSSLVLTIIIDVTCVGNMEKNRLDNTNQNLYHHNP
EEVCSFCNCSERIDVDADVVDVVELLYVSENFNGEDSAIGTVTIGVEVEAEMLR
"

complement(4667..6566)

/gene="F2009.30"

/number=1

4942..4986

/gene="F2009.40"

/number=1

4987..5044

/gene="F2009.40"

/number=2

5045..5095

/gene="F2009.40"

/number=2

5096..6188

/gene="F2009.40"

/number=3

complement(6567..6669)

/gene="F2009.30"

/number=2

complement(6670..6877)

/gene="F2009.30"

/number=2

complement(6878..7149)

/gene="F2009.30"

/number=3

complement(7150..7304)

/gene="F2009.30"

/number=3

complement(7305..7371)

/gene="F2009.30"

/number=4

complement(7372..7536)

/gene="F2009.30"

misc_feature	complement(7930..8000) /note="exon predicted by xgrail, quality good_shadowexon" 8174..8302
misc_feature	/note="exon predicted by xgrail, quality excellent" complement(8465..8586)
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 8544..8629
misc_feature	/note="exon predicted by xgrail, quality good" 8702..8829
misc_feature	/note="exon predicted by xgrail, quality good" complement(8934..9057)
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 8936..9051
misc_feature	/note="exon predicted by xgrail, quality excellent" 9147..9243
misc_feature	/note="exon predicted by xgrail, quality excellent" 9290..9338
misc_feature	/note="exon predicted by xgrail, quality marginal" complement(9317..9461)
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 9492..9555
misc_feature	/note="exon predicted by xgrail, quality marginal" complement(9654..9780)
misc_feature	/note="exon predicted by xgrail, quality marginal_shadowexon" 9866..9814
misc_feature	/note="exon predicted by xgrail, quality excellent" complement(9912..9988)
misc_feature	/note="exon predicted by xgrail, quality marginal_shadowexon" join(<10040..10154,10340..10507,10608..10936, 11019..>11237) /gene="F23N11.3" <10040..>11237 /gene="F23N11.3" /gene="F23N11.3"
mrna	/note="F5H14.37; predicted by genscan" join(10040..10154,10340..10507,10608..10936,11019..11237)
gene	/gene="F23N11.3" /note="hypothetical protein" /codon_start=1 /protein_id="AAD21704.1" /db_xref="PID:g4512649" /db_xref="GI:4512649" /translation="MTLLVYKGAARIVSMIGQAAIMDSYLCLLHLAGILVESLFNA
CDS	

Query Match	7.0%	Score 134.2;	DB 8;	Length 61384;
Best Local Similarity	63.7%;	Pred. No. 2.1e-08;		
Matches 221;	Conservative 0;	Mismatches 123;	Indels 3;	Gaps 1;
QY 724	agcgcctggcagagcaccctgcgcagggggtgcgcgtacatgctatcgccgcagcagcgtc	783		
Db 35491	AGAGCCGTGCTGAATTATATGAACCGAGATCTCCGCACAAAGTTAACGCCTGAAGATATT	35550		
QY 784	ttctcaccgcgcgggagccagcgcatcgagtgctcataatcccggtgctggcccgacc	843		
Db 35551	TTTCTGACCGCTGGATGCACACCAAGGATAGAGATCGTTCGAATCGTGGCTCGACCA	35610		
QY 844	gcgcggcgcaacattctgtctccccagggccaggtctacccaactacgagggcgcgccgcgcg	903		
Db 35611	AA---CGCAAAATCTTGCTCCCAAGCTCTGGCTTCCCTCACTACGACGCTCGTGTGCT	35667		
QY 904	ttcaacagcgtgcaggtccgcgcatttcgacctcatccccacaaaggggtggagatcgac	963		
Db 35668	TACAGTGTCTCGAGGTTCCGAAGTTTGATCTCTCCCGAGAAAGAAATGGGAGATTGAT	35727		
QY 964	atcgactcgtggaattccatccgcgcacgaacaccaccgcctatggtcatcataaacccc	1023		
Db 35728	CTTGAAGGTATCGAAGCCATTGCAGACGAGACACTTGGCTATGGTGTGAATTAAACCC	35787		
QY 1024	aaacaccgcgtgcgcagcgtttactctcagaccatctgtccaaagt	1070		
Db 35788	AACAATCCCTGTGGAATGCTACTCTCAGGACCATCTCAAAAGGT	35834		

```

RESULT 5
RATTATRA
LOCUS      RATTATRA             2004 bp      mRNA           ROD           15-MAR-1990
DEFINITION Rat tyrosine aminotransferase mRNA, complete cds.
ACCESSION M8340 K01265
VERSION   M8340.1 GI:207155
KEYWORDS  tyrosine aminotransferase.
SOURCE    Rat (adult Wistar) hepatocyte, cDNA to mRNA, clones pCTAT[1-4].
ORGANISM  Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
          1 (bases 1 to 2004)
          Hargrove,J.L., Scoble,H.A., Mathews,W.R., Baumstark,B.R. and
          Blemann,K.
TITLE     The structure of tyrosine aminotransferase: Evidence for domains
          involved in catalysis and enzyme turnover
JOURNAL   J. Biol. Chem. 264, 45-53 (1989)
MEDLINE   89079691
COMMENT   Draft entry and computer-readable sequence for [1] kindly submitted
          by J.L.Hargrove, 07-NOV-88.
FEATURES  source
          location/Qualifiers
          1..2004
          /organism="Rattus norvegicus"
          /db_xref="taxon:10116"
          98..1462
          /note="tyrosine aminotransferase (EC 2.6.1.5)"
          /codon_start=1
          /protein_id="AAA42203.1"
          /db_xref="PID:g207156"
          /db_xref="GI:207156"
          /translation="MDSYVIVTQDSDLSSVLVDVHVNIGNSVQCKRKGKRWDRV
          PDSMSTNPIRAKTDNNKVPNPKTVISLIGDPTVEGNLPTDPEVTOAMKDAL
          SKYNGVAPSIGVLSREVASVYHCHAEPLAKDVLISGGSQAIECLAVLNPQG
          NLLIPRGSFLTKLAESMGIEVKLLNLLPEKSWEDLKQLESLIDKEKLECLVNNPS
          NPGCVSKRHKQILAVAEKQVPIADEIYEDVMVFSDCKVEPLANLSTNVPILSCG
          POBFWHTLSFKSNADLCYALADLPQLPQPSGAMVLMVGIEMHFEFENDVEF
          TERLIAEQAVHCLPATCFEVPNFVITVPEVMMLEACSRIQEFCEQHYHCDGSGQE
          ECDK"
BASE COUNT      491 a      521 c      517 g      475 t
ORIGIN

Query Match      6.9%; Score 131; DB 12; Length 2004;
Best Local Similarity 49.1%; Pred. No. 1e-07;
Matches 493; Conservative 0; Mismatches 500; Indels 12; Gaps 5;

QY 549 cagcgagcgtgcagagagaagggcgccgctgtcgtcgcctgtgcccacgggacc 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 CAACATGAAGTGCAGGCCCAATCCGACAAAGACCGTGATTCTCTGTCAATTGGGACC 339

QY 609 gtccgtgtcccgcccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 TACTGTGTTGGNACCTGCTACAGNCCCTGAAGTACCACAGCCATGAAGATGCCCT 399

QY 669 gcgcaccggccagttcaactgtaccgccgcgcgcgcgcgcgcgcgcgcgcgcgc 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 GGACTCGGGGAAGTACAATGGCTATGCCCGTGCATCCGGTACCTATCCAGTCGGGAGGA 459

QY 729 cgtggcagagcacctgtcgcagggcgtgcgtacatgctatcgccgcgcgcgcgc 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 GGTGCGTCTTAC---TACCACCTGTGATGAGGCTCCTCTGGAAGCTAAGGATGCTATCT 516

QY 789 caccgcgcggggaccaggcgcgtcagggttcataatacccggtgctggcccagac 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 GACAAAGCGGTGCAGTCAGGCCATTGAGCTATGCTAGTGTGTTGGCC---AATCCTGG 573

QY 849 cgccaacattctgctccccaggccaggtaccaccaactacgagggcgcgccgcttaa 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 ACAAAACATCCTCATTCCAAGGCCGGGTTTCCCTCTATAGACTTTGGCTGAGTCTAT 633

```



```

QY 1151 tcatccaatggaggtgtt---tgggcacatcacccctgtgctgcataggggtctctgt 1207
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33865 ACAATCCCTGGCGCTCTCCCGGGGATGCGGGAGCAGCAGGTGACCATCGGTGCGCCG 33924

QY 1208 ccaagtcatgtagtgcctggatgcggcgttggtgtagggtagcgg 1252
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33925 GCAAGAGCTTCTGTTCAACGGGCTGGAAGGTGGCTGGGTCAACG 33969

RESULT 7
HSTATR      2754 bp      mRNA      PRI      12-SEP-1993
LOCUS      Human mRNA for tyrosine aminotransferase (TAT) (EC 2.6.1.5).
DEFINITION
ACCESSION  X52520
NID         936712
VERSION     X52520.1 GI:36712
KEYWORDS    aminotransferase; transferase; tyrosine aminotransferase.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2754)
AUTHORS     Scherer,G.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAR-1990) Scherer G., Institute of Human Genetics,
            Albertstr 11, D 7800 Freiburg
REMARK      9bases 1 to 2754)
REFERENCE   2 (bases 1 to 2754)
AUTHORS     Rettenmeier,R., Natt,E., Zentgraf,H. and Scherer,G.
TITLE       Isolation and characterization of the human tyrosine
            aminotransferase gene
JOURNAL     Nucleic Acids Res. 18 (13), 3853-3861 (1990)
MEDLINE     90326506
COMMENT     See X52507>-<X52520>.
            Data kindly reviewed (26-JUL-1990) by G. Scherer.
FEATURES    source
            Location/Qualifiers
              1..2754
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="16q22.1"
                /tissue_type="placenta (genomic) and liver (cDNA)"
                /clone_lib="genomic (phage) + cDNA"
                /clone="lambda-hvAT1, phcTAT2-16 + phcTAT3a-6"
                97..1461
                  /note="tyrosine aminotransferase (AA 1-454)"
                  /codon_start=1
                  /protein_id="CAA36750.1"
                  /db_xref="PID:g36713"
                  /db_xref="GI:36713"
                  /translaton="MDPYMIQMSKGNLPSILDVHVNVGSRSSVPGKMKGRKARWSVR
PSDMAKTFNPRAIVONMKVPKPNKTMISLSIGDPTVFGNLPDPEVTAQMKDALD
SGKYNVATPSIGFLSSREETAASYHCPEAPLEAKDVLTSCQSIDLCVLANPQO
NILYPRGFSLYKTLAESMGIEVKLYNLPKLEWEIDLKOLEYLDKETAQLVNPQS
NPCGSFSKRLQKILAVARQCVPILADEHYIGDMVSDCKYEPLATLSDVILSCG
GLAKRWLPWGRILGWILIHDRDIFGEINRDLGVLKLSQRIIGPCTIVQGALKSLCET
PGEYHTLFLKSNADLCYKALAAIEPLRPVSPGAMVLMVMEHHPFEFENDVEF
TERLVAEQSVHCLPATCFEYPNFIRVITVPEVMMLEACRSRIOEFCEQHYHCAEGSOE
ECDSK"
              2070..2075
                /note="minor mRNA polyadenylation signal"
              2090
                /note="minor mRNA polyadenylation site"
              2116..2130
                /note="direct repeat flanking Alu element"
              complement(2131..2479)
                /note="Alu element"
              2480..2494
                /note="direct repeat flanking Alu element"
              2738..2743
                /note="major mRNA polyadenylation signal"
              2754
                /note="major mRNA polyadenylation site"

```

```

BASE COUNT      689 a      646 c      635 g      784 t
ORIGIN

Query Match      6.2%; Score 118; DB 10; Length 2754;
Best Local Similarity 50.1%; Pred. No. 3.3e-06;
Matches 348; Conservative 0; Mismatches 340; Indels 6; Gaps 2;

QY 581 csgtgcgcgtgcccacggggaccgcctcggtgttcccgccctcccgacggcgctcg 640
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 CCATGATTTCCCTGCTCCATTTGGGACCTACTGTGTTTGAACCTGCCCTACAGACCTG 370
QY 641 agccgagagacgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 700
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 AAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGGCAAAATATAATGGCTATGCCCAT 430
QY 701 gcctggcctccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 760
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 CCATCGGCTTCTTATCCAGTCGGGAGGAGATGCTCTTTA---TTACCACTGTCTGAGG 487
QY 761 acatgctatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 820
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 488 CACCCCTAGNAGCTAAGGAGCTCATTTGCAAGTGGCTGCAGCCAAGCTATTGACCTTT 547
QY 821 taatcccggtgtgcccacggaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 880
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 548 GTTTAGCTGTGTGGCCAA---CCAGGGCAGAACATCTCTGGTTCCAAGACCTGGTTCT 604
QY 881 caaatcacgagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 940
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 605 CTCTTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACTCTAGCAATTTGTTGC 664
QY 941 ccgacaaggggtggagatcgacatcgactcgtggaatccatcgccgacagaacacca 1000
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 CAGAGAAATCTTGGAAATTTGACCTGAAACAACCTGGAATATCTAATTGATGAAAGACAG 724
QY 1001 ccgcatggttcataaaccacacaccccgcgccgcgcgcgcgcgcgcgcgcgcgcgcgc 1060
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 725 CTTGTCTCATTTGCAATAATCCATCAACCCCTGTGGTCAGTGTTCAGCAACGTCATC 784
QY 1061 tgcctcaagtcgcggaggtgsgaaaggtcgcgaatattggtgattggaagaggtat 1120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 785 TTCAAGACATCTTGGCAGTGGCTGCAGCGGACGTGTGCCCATCTTAGCTAGATGATCT 844
QY 1121 acggcaagcgtgttctggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 845 ATGGAGACATGTTGTTTTCGGATTGCAATATGAACCACTGGCCACCCCTCAGCACCGATG 904
QY 1181 cccctgctgtccataggtctctgtccaaagtcagtagatagtcctggatggcgcttg 1240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 905 TCCCATCTCTGCTGTGGAGGGCTGGCCAAAGCGCTGGCTGGTTCCTGGCTGGAGGTGG 964
QY 1241 gatggtagcgggtgtacaccccgacagaagaatctt 1274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 965 GCTGATCCCTCATTCATGACCGAGAGACATTT 998

RESULT 8
HSTYAT      2051 bp      mRNA      PRI      13-JUL-1995
LOCUS      H. sapiens mRNA for tyrosine aminotransferase.
DEFINITION
ACCESSION  X55675
NID         937501
VERSION     X55675.1 GI:37501
KEYWORDS    tyrosine aminotransferase.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2051)
AUTHORS     Labrie,F.
TITLE       Direct Submission
JOURNAL     Submitted (26-SEP-1990) Labrie F., CHUL Research Centre, Laboratory

```


Db 9569 GACGTACCGGGTGGGGTTCCGCTCCGGGTGGCTGGTGGTGCACGGGGCCGAAGCAG 9512

RESULT 11
MSGB1970CS/c
LOCUS MSGB1970CS 39399 bp DNA BCT 26-JUN-1996
DEFINITION Mycobacterium leprae cosmid B1970 DNA sequence.
ACCESSION L78815
NID g1377775
VERSION L78815.1 GI:1377775
KEYWORDS
SOURCE
Mycobacterium leprae (clone: cosmid B1970) (tissue library: Lorist 6) DNA
ORGANISM Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE
1 (sites)
Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
2 (bases 1 to 39399)
Smith,D.R., Richterich,P., Rubenfield,M., Butler,C., Lee,H.-M.,
Xu,Q., Gunderson,K., Chung,M., Maher,J.K., Deloughery,C.,
Aldrich,T., Imrich,J., Tulig,C., Smyth,A., Drill,S., Avruch,A.S.,
Rice,P., Abendschan,K., McDougall,S., Deloughery,C., Kirst,S.,
Safer,H., Connolly,S., Abendschan,K., Eiglmeier,K., Bergh,S.,
Cole,S., Robison,K., Jaehn,L., Gryan,G., Church,G.M. and Mao,J.
Prepublication submission
Unpublished (1996)
TITLE
JOURNAL
COMMENT
This sequence was generated by the Genome Sequencing Center at
Genome Therapeutics Corporation (Collaborative Research Division),
100
Beaver St., Waltham, MA, 02154. Please contact Doug Smith
(smith@eric.com) for further information. The sequence represents
the
insert of a Lorist 6 cosmid clone from a mapped set of clones
constructed from M. leprae genomic DNA isolated from armadillo
liver
[3]. The sequence may not represent the entire cloned insert of
the
cosmid if an overlapping region was previously sequenced from
another
clone. Coding sequences larger than 60 amino acids were predicted
on
the basis of codon usage and homology information. An attempt was
made
to locate the most probable start site based on codon usage,
homology,
the presence of a Shine-Dalgarno sequence, or overlapping orf that
suggested translational coupling. It is possible that the actual
start
site differs from the one selected.
FEATURES
source
1..39399
Location/Qualifiers
/organism="Mycobacterium leprae"
/specific_host="Dasypus novemcinctus"
/db_xref="taxon:1769"
/clone="cosmid B1970"
/tissue_lib="Lorist 6"
/note="The liver of the armadillo was used to isolate the
Mycobacterium leprae."
BASE COUNT 7827 a 10529 c 12083 g 8960 t
ORIGIN
Query Match 5.8%; Score 111; DB 1; Length 39399;
Best Local Similarity 51.8%; Pred. NO. 1.3e-05;
Matches 300; Conservative 0; Mismatches 275; Indels 4; Gaps 2;
Oy 689 gctaccgccggcggtggcctcccccgcgcagacgcgcgtggcagagcactctgcg 748

[illegible]

Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
NATURE 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]
2 (bases 1 to 9150)
Direct Submission
Parkhill,J.
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2276330.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of rB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
Location/Qualifiers
1..9150
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1..9106
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y279"
82..86
/note="possible RBS upstream of Rv0336"
90..1601
/gene="Rv0336"
90..1601
/gene="Rv0336"
/note="Rv0336, (MTCY279.03), len: 503 aa: almost identical to M. tuberculosis hypothetical protein MTCY20610.05 (99.8% identity in 503 aa overlap) possibly due to a recent gene duplication, also similar to other M. tuberculosis hypothetical proteins eg. MTCY049.05 (61.1% identity in 347 aa overlap) and Y140_MTCYU Q10709 hypothetical 58.9 kd protein cy49.40 (550 aa), fasta scores: Opt: 676 z-score: 978.7 E():0.41.7% identity in 513 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0336"
/protein_id="CAB10791.1"
/db_xref="PID:e330320"
/db_xref="PID:e330320"
/db_xref="GI:2276333"
/db_xref="GI:2276333"
/db_xref="SPTREMBL:O33266"
/translation="MPSPEAIAHDFEFCHAPKTRTVSAAFIDRICASRAENRAAA AQLVAGELFAIRMSRCGGREWMYDTHAAVEAAALRSISQGLAASRLRYARAME RLKPTAVFSGADIGYLMFATIVYRTDILVDPDLAAVDAQLAANVARWPSMTKARLA GQNDKIYARADADAVRREKYOQORQFWGESODGVCQIGSLIADVDAHALDARLSAL AGTVCEDHPSREORADALGALAGGADRLCGCGCAGACGACRAKRPAPPPVTHLIAEA ATINGTSPASQNMNADLITAEILVAELAKTATLVLVHPGDAPPEPGYAPSPKALADF VCRDLTCRWFQENCDLDTIFYAAGGPTTHSLKCYCRTHLVLTFWGRDQ QLPDGLTILTSPSGHTVYVTPGSALEPFLPSLCHFSGGIPAPEADPPYDHCQRTAMPEK

RRRTAQRDAYRIATERRONHAAQRAQVLTOTAAATDTHGPPPDNDPPPF"
complement(1771..3060)
/gene="aspC"
complement(1771..3060)
/gene="aspC"
/note="Rv0337c, (MTCY279.04c), len: 429, aspC, aspartate aminotransferase, similar to many eg. TR:Q48143 (EMBL:U32823) aspartate aminotransferase (404 aa), fasta scores: Opt: 1646 z-score: 2132.9 E(): 0, 57.2% identity in 404 aa overlap; also some similarity to M. tuberculosis protein MTCY06G11.12 (27.2% identity in 383 aa overlap)"
/codon_start=1
/transl_table=11
/product="aspC"
/protein_id="CAB10792.1"
/db_xref="PID:e330321"
/db_xref="PID:g2276334"
/db_xref="GI:2276334"
/db_xref="SPTREMBL:O33267"
/translation="MNDGTIVDTVTHQLPWHTASHORAFQAQSAKLQDLVYEIRGP VHOAAARLAEGRHILKLNIGNPAFGFEAPDIVMRDIIQALPYAQYSDSQILSAR RAVTVYELVPGPRFDVDDVYLGNGVSELTITMTLOALDNGDQVLIPSPDYPLMTAS TSLAGCTPVYLDDETQGPQDIDLESKITERTKALVINPNPNTGAVYSCIELTQM VDAKQHQLLLADEIDIKLYDDAKHISLASIAPMCLTIFNGLSKAYNAVYRAGW LATITPKHASSFIEGILLANMRLCPNVPQAIOVALGHQSIEDLVLPGRLLQEQ RDTAMTKLNEIPGVSCVKGALYAFPLRDPDYDIDDDQLVDLLLSKILVTQGT GFNWPADHLRLVTLFWSRDLAAAIERLGNFLVSRQ"
complement(3091..5739)
/gene="Rv0338c"
complement(3091..5739)
/gene="Rv0338c"
/note="Rv0338c, (MTCY279.05c), len: 882, unknown iron-sulfur protein, first of several possible start sites chosen. Some similarity to other hypothetical proteins eg. YWJF_BACSU P45866 hypothetical 79.2 kd protein in acda (705 aa), fasta scores: Opt: 1188 z-score: 894.6 E(): 0, 34.8% identity in 682 aa overlap. Contains PS00070 Aldehyde dehydrogenase:cysteine active site and two of PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0338c"
/protein_id="CAB10793.1"
/db_xref="PID:e330322"
/db_xref="PID:g2276335"
/db_xref="GI:2276335"
/db_xref="SPTREMBL:O33268"
/translation="MTQTILRLILGSMATVGVFALRVWMLYKLVMSGQPSAGRT DNLGTRITQISEVLGQRLLKWSIFGLAHFTTMMGFFILLTVIYAYGLLFEERFI PVTRWDALGLODFATVFLGITTFAIRILRNPRNREIGSRFYSHGNGAWLVLL MIVNVIWTVYLVRSVAVNGTLPYNGAFSLQSLFALRPLGQNPANIEITALLHLL GVMNLELILVHSHLHLFLAPINVTFKLPDGLPPLPLEADGKPLDFENPSDAVE GKRKIDFTWKGMLDFATCECRCSQCPANWTGKPLSLVINDLRDHWMAKAPYL LGQDSAGGAGHGEHHVPSGFGVPGHPEQATRLPLVLEEGGVIDDPDVLWSC VTCGACVEQCPDIEHYDVMRRYQWMMSEFSELSLFLKNETKGNPQGNASD RTNIDVEDFDVPVYQGVDFDGYEYLFVWGCAGAYDDAKKTTKAVALLAVARYK VYLIGAGETCNGDSARSGNEFLFOOLAQAVETLDGLFEGVETVDRKIVVTCPCFN TICKEYRQLGANTYVLHTOLLNRLVRDKELVPTVPVSODITYVHPCYLGHNKAYEA PRELIGAAGASLTMPRHADRSFCCGAGGARMWEEHIGRINHERVDEALDATAI ATACPFRCVMVTDGVNDREBEAGSVEVLDAVLLGSLDHDKAUPAGTKAAQAO ERAPKAAPKAAAPVTPVEAPAPAPAPAPAPAPVKGGLGMAAKRPPGAKKAPTPA AAPAAPAPVGLGTAAGKRPAGKTPPPAPGLAEPAAQPEAKPQPEAPAPKPT DGDPAAPAPVKGIGTARGARPPGKR"
complement(4174..4209)
/gene="Rv0338c"
/note="PS00070 Aldehyde dehydrogenases cysteine active site"
complement(4531..4566)
/gene="Rv0338c"
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature"
complement(4822..4857)
/gene="Rv0338c"

misc_feature

misc_feature

misc_feature


```

CDS
/ gene="stgR"
/ 2219..3202
/ gene="stgR"
/ note="lysR-type transcriptional regulator"
/ codon_start=1
/ transl_table=11
/ product="StgR protein"
/ protein_id="CAA06437.1"
/ db_xref="PID:el1318409"
/ db_xref="PID:g3550624"
/ db_xref="GI:3550624"
/ db_xref="SPTREMBL:O88132"
/ translation="MDLDLALVRSFAITAEELHFRRAERLFTIOGLSQRIRLELL
LGVSLFVNGNRVTEGQFLAPARRVMEADAADAAVQVPRRRPLRVVDGRLA
PLQVLRKLEEDPELPLNSMROSLAALPALARDEIDTAFGRVHDVQGWAEFSHRL
VRLEPLRALVPRTHPLAGRDVRIDELRSALWQPRNGVPSEWEHYVGRADSFGLSL
AFSGAALSDHEFLEHLNHWGEMACLAGADVSTAFSPDIRSLPLVDPTFVYPMWVWRR
QAGHPLVDRLVGLAKRSAGDWLTHVPGELWLPAPDRALLRGAGVDVDAVRG"
BASE COUNT      447 a 1112 c 1189 g 492 t
ORIGIN

Query Match      5.4%; Score 102.2; DB 1; Length 3240;
Best Local Similarity 50.9%; Pred. No. 0.00024;
Matches 301; Conservative 0; Mismatches 278; Indels 12; Gaps 2;

QY 679 cagttcaactgtaccgcccgccggtcggcctcccccgcgcacgaagcgcgctggcgag 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1202 CAGGCGCAGCGGTACACGACTCGCGCGCGCTCGTCTCCGCCGCCGCCGCCGCCAG 1143

QY 739 cactgtcgaggcggtgcgtacatgctatcgccgcgacgacttctctaccgccggc 798
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1142 CGTATACAGCGCTCGGCTCGACGAGGTCTCGGTGGACGAGCTTCTCTCGGCAACGCG 1083

QY 799 gggaccagcgatcgagtcataatcccggtgctggcccgacacgcgcggcgcccaactt 858
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1082 GTGTCGAGCTGTCAGATGCGGTGTCAGGCGCT---CCTGGAGGACGGCGACGAAATC 1026

QY 859 ctgtcccccagcgccggtaccacaaactacgagcgcgccgcttcaacaggctggag 918
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 CTCATCCCGCGCGCGACTTCCCGCTGTGGACCGGGTGACACCTCTCCGGAGGCAAG 966

QY 919 gtccggcatttcgaacctatccccacaaaggggtgggagatcgacatcgactcgtggaa 978
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 GCCGTCCACTAGCTGTGCGACGAGCGCGGACTGCTACCCGACCTCGACGACATGGCC 906

QY 979 tccatcgccgacagaacacaccgccatggtcatcataaaccacacaccccgctgcggc 1038
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 TCGAAGATCACCACACGACCAAGCCGTCGAGATCATCAACCCCA-----CGGGC 855

QY 1039 agcgtttactctacgacctctccaaggtcgcgaggtggcgaaaggctcggaata 1098
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 GGGGTCTATCCGAGAGATCTGGAGGCACTCTCGATCTCGCGCGGCGGACGGCCTG 795

QY 1099 ttgggtattgtcgaggtatatacgaggaagtgttctgggagcgcccgcttcatccca 1158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 ATGGGTCTTCGCGGACGAGATCTACGACCAATCTGTACGAGCGCGCTGCACCAACAGC 735

QY 1159 atgggaggtttgggcacatcacccctgtgtccatagggtctctgtccaagtcatgg 1218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 GTGGCGGCGCTCGCCCGGATCTGTGGTCCCTACCTTCTGCGGTCTTTTGAAGACGTAC 675

QY 1219 atagtgcctggatggcggttgatgggttagcggtgtacgacccccagaaag 1269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 CGCGTGGCGGTTTCGGTTCGGGCTGGTGGTTCATCACCGGCGCGCGGCGAG 624
```


Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	DB			
1	c	1910	100.0	1910	1	V48148	Nicotianamine amin	
2		1146.4	60.0	1660	1	V48147	Nicotianamine amin	
3		95.4	5.0	390	1	Q28833	Randomising oligon	
4		95.4	5.0	390	1	Q36859	PCR primer for 5'	
5		80.2	4.2	985	1	V44439	Mycobacterium tube	
6	c	80.2	4.2	985	1	V64548	M. tuberculosis im	
7		77.8	4.1	3946	1	T93610	Mycobacterium tube	
8		75.4	3.9	2539	1	T61454	Streptomyces virid	
9		75.4	3.9	2809	1	T61455	DhP4-mel chimeric p	
10		74.2	3.9	2277	1	V05372	Human telomerase p	
11	c	74.2	3.9	2277	1	V13836	Human telomerase p	
12		73	3.8	114955	1	X53491	Homo sapiens mamma	
13		69.6	3.6	114955	1	X53491	Human adenosine A1	
14		69.4	3.6	3946	1	T93610	Human adenosine A1	
15		67.4	3.5	44377	1	T78508	Mycobacterium tube	
16	c	67.4	3.5	44377	1	T80414	Platenolide syntha	
17		67.2	3.5	1439	1	V44424	Mycobacterium tube	
18		67.2	3.5	1439	1	V64533	M. tuberculosis im	
19		66.6	3.5	15872	1	T68715	Streptomyces venez	
20		65.8	3.4	1620	1	Q22482	groEL-1 gene codin	
21	c	65.8	3.4	2167	1	Q22484	groEL-1 gene. Reco	
22		65.8	3.4	2668	1	Q22485	groEL-1 gene. Reco	
23		65.2	3.4	2742	1	V49291	Aufjesky's disease	
24		65.2	3.4	1166	1	X34248	Mycobacterium spec	
25		65.2	3.4	1165	1	X34249	Mycobacterium spec	
26	c	65.2	3.4	1164	1	X34250	Mycobacterium spec	
27		64.2	3.4	2198	1	T71711	Cellulomonas fimi	
28		63.8	3.3	24379	1	T93095	Streptomyces freno	
29		63.8	3.3	24379	1	V25925	Streptomyces roseo	
30		63.2	3.3	494	1	V44433	Mycobacterium tube	
31	c	63.2	3.3	494	1	V64542	M. tuberculosis im	
32		62.8	3.3	1272	1	X09010	Bpr-3a polynucleot	
33		62.6	3.3	15672	1	Q10613	Rianodin receptor	
34		62.6	3.3	30001	1	T61016	Total DNA sequence	
35		62.6	3.3	985	1	V44439	Mycobacterium tube	
36	c	62.6	3.3	985	1	V64548	M. tuberculosis im	
37		62.6	3.3	30001	1	X05110	S. aureofaciens DN	
38		62.6	3.3	538	1	X34437	Mycobacterium tube	
39		62.2	3.3	538	1	V64546	M. tuberculosis im	
40		62	3.2	1320	1	Q22481	groEL-1 gene parti	
41	c	61.8	3.2	29879	1	Q46806	eryA region of S.	
42		61.8	3.2	2888	1	T93268	Streptomyces prist	
43		61.2	3.2	12001	1	Q76213	HSV L/ST region. H	


```

QY 255 gaacggcaagagcaacggccatgccaggccactgcgaacgagcgaacggcgagccactgc 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 TGGCGCGCGCGCACAG---TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 315 gaacggcaagacaaacggccacgcgagagcaacgagccatgctgagcccgccgagcgaa 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TACCGCGCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
QY 375 cggcgagagcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
QY 428 ggcacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TGACGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
QY 488 aggcacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
QY 548 tcagcgcgagcgtgcagggagaaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 AGGTCTCGCGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
QY 608 cgtccgtgttcccgcccttcgcacgcgctgcagcgcgcgcgcgcgcgcgcgcgcgcgcg 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACGGGCGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 668 tgcgcacggcagcttcaactgtaccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ATGGCGGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639

RESULT 6
V64548
ID V64548 standard; DNA; 985 BP.
AC V64548.
DT 27-JAN-1999 (first entry)
DE M. tuberculosis immunogenic polypeptide XP25 3'-end DNA.
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
   vaccine; pharmaceutical; infection; diagnosis; ss.
OS Mycobacterium tuberculosis.
PN WO9816646-A2.
PD 23-APR-1998.
PF 07-OCT-1997; U18293.
PR 13-MAR-1997; US-818112.
PR 11-OCT-1996; US-730510.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SG, Skeiky IAW, Twardzik DR, Vedvick TS;
DR WPI; 98-261042/23.
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
PS Claim 4; Page 176; 230pp; English.
CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC tuberculosis (MT) antigen which can be used in a method for inducing
CC protective immunity against tuberculosis (TB). This sequence can be
CC formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match 4.2%; Score 80.2; DB 1; Length 985;
Best Local Similarity 48.7%; Pred. No. 8.9e-07;
Matches 312; Conservative 0; Mismatches 318; Indels 11; Gaps 3;

QY 75 catggccaccgtacgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 CAGCGGTACCGGTGGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62
QY 135 cgcgaacggcaagagcaacggccatgctggtgctggtgctggtgctggtgctggtgctggt 194

```

```

Db 63 CAACACGGGTGTGGGGGGACCAACCGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCG 122
QY 195 ccatggtgtggtatgccagcgcgagcaacgagcaacgagcaacgagcagcagcagcagc 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 CGCACCGGAGGTGCGGGGGGTCCGGCGCGGACAAACCCACCGGTGCTGTTTTCGCCGG 182
QY 255 gaacgcaagagcaacgagccatgctcagagccactgcgaacgagcagcagcagcagcagc 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 315 gaacgcaagagcaacgagccacgcgagagcaacgagcagcagcagcagcagcagcagc 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TACCGCGCGCACCGCGCGCGTTTGTCCGCGCCACCGGTAGTGACAGGATCGCGCGGCGCG 299
QY 375 cggcgagagcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CGCGCGCGCGGTGACGGCGCGATGGGCGCAGCGGTCTCGGCTCGGCTCTCCGGCTT 359
QY 428 ggcacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TGACGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
QY 488 aggcacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
QY 548 tcagcgcgagcgtgcagggagaaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 AGGTCTCGCGACACGGCGCGCGGTTCGCGCGGTGACGGTGGGCGCGGTGGCGC-CGCCGGCA 538
QY 608 cgtccgtgttcccgcccttcgcacgcgctgcagcgcgcgcgcgcgcgcgcgcgcgcgcg 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACGGGCGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 668 tgcgcacggcagcttcaactgtaccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ATGGCGGCAACGGCGCGCGCGCGCGCGGTGCTGGCGGGCGCGCGCGCGCGCGCGCG 639

RESULT 7
T93610/c
ID T93610 standard; DNA; 3946 BP.
AC T93610;
DT 27-APR-1998 (first entry)
DE Mycobacterium tuberculosis genomic DNA fragment (I).
KW Tuberculosis; mycobacteria; infection; diagnosis;
   antimycobacterial; antibiotic; vaccine; ss.
OS Mycobacterium tuberculosis.
PN WO9741252-A2.
PD 06-NOV-1997.
PF 18-APR-1997; E01973.
PR 29-APR-1996; DE-017184.
PA (GBFB ) GBF GES BIOTEC FORSCHUNG GMBH.
PI Espitia C, Honisch C, Moreno C, Singh M;
PI WPI; 97-549750/50.
DR P-PSDB; W31852-55.
DR New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
PT Claim 1; Fig 9; 55pp; English.
PS New claimed DNA (A) has one of 3 isolated Mycobacterium
CC tuberculosis DNA sequences of 3946 bp (I) (T93610), 2653 bp (VI)
CC (see T93611) or 440 bp (IX) (see T93612), or hybridises to, or is a
CC fragment of these sequences. Also claimed are RNA transcripts
CC of (A), and proteins encoded by (A). To isolate (I), a cosmid
CC library of M. tuberculosis DNA was screened with degenerate probes
CC (see T93621) containing GC-rich regions and designed to encode part
CC of a proline-rich protein. (I) contains open reading frames (ORFs)
CC for 4 proteins (see W31852-55). These ORFs can be amplified and
CC cloned into vectors for expression in E. coli or other hosts.
CC Clones (VI) and (IX) were obtained from a phage library and
CC together encode 3 proteins (see W31851 and W31856-57). The claimed

```


RESULT 10
ID V05372 standard; RNA; 2277 BP.

AC	V05372;
AT	O6-JUL-1998 (first entry)
DE	Human telomerase p105 subunit mammalian optimised synthetic RNA.
KW	Telomerase; p105; human; cell replication; cancer; restenosis;
KW	multiple sclerosis; inflammation; rheumatoid arthritis;
KW	myocardial infarction; glomerulonephritis; transplant rejection;
KW	infection; therapy; ds.
OS	Homo sapiens.
OS	Synthetic.
PN	WO9801542-A1.
PD	15-JAN-1998.
PF	08-JUL-1997; U12296.
PR	08-JUL-1996; US-676974.
PA	(REGC) UNIV CALIFORNIA.
PI	Collins K;
PI	WPI; 98-101043/09.
PT	New nucleic acid encoding human telomerase proteins or their
PT	fragments - useful for therapeutic modulation of telomerase activity
PT	and for screening for potential modulators of telomerase-target
PT	binding
PS	Disclosure; Page 23-24; 32pp; English.
CC	This polynucleotide comprises a synthetic RNA sequence coding for
CC	the p105 subunit (see W46593) of human telomerase and optimised for
CC	expression in mammalian cells. It is based on an isolated cDNA
CC	clone (see V05369) for p105. The invention provides methods relating
CC	to human telomerase and related nucleic acids, including the subunit
CC	proteins p140, p105, p48 and p43. The proteins may be produced
CC	recombinantly from transformed host cells or purified from human
CC	cells. Also included are human telomerase RNA (see V05373) and
CC	functional derivatives (see V05374 and V16092-93), as well as p105
CC	synthetic DNA sequences (V05370-72). The invention also provides
CC	isolated telomerase hybridisation probes and primers capable of
CC	specifically hybridising with the telomerase gene, telomerase-
CC	specific binding agents such as specific antibodies, and methods of
CC	making and using the subject compositions in diagnosis (e.g.
CC	genetic hybridisation screens for telomerase transcripts), therapy
CC	(e.g. gene therapy to modulate telomerase gene expression) and in
CC	the biopharmaceutical industry (e.g. reagents for screening
CC	chemical libraries for lead agents). Modulation of telomerase
CC	expression can be used for the treatment or prevention of cancer,
CC	restenosis, inflammation, myocardial infarction, glomerulonephritis,
CC	cancer, transplant rejection or infections (e.g. with HIV).
CC	Sequence 2277 BP; 557 A; 723 C; 781 G; 216 T;
SQ	

Query Match 3.9%; Score 74.2; DB 1; Length 2277;
Best Local Similarity 45.8%; Pred. No. 1.3e-05;
Matches 333; Conservative 0; Mismatches 388; Indels 6; Gaps 2;

QY	75	catggtccatcgctacgcagcagcaggagtccgccgaacggccttgcgtggcgcgacg	134
DB	585	CAGGACACCAGCGCTGAGGCCCATCGCGAGGAGAACGACGCACGAGCAAGCACCA	644
QY	135	cgcgaacgcgaagagcaacggccatggcgtgcgcgcgtgaacgcgcaagacaacgg	194
DB	645	GGAGAGCGTGAAGAAGAAGCGCGCAGGAGGAGGACATCGGAGGAGGAGGACGACGA	704
QY	195	ccatggcgtggatcgcgacgcgaacggccaagagcaacggccatggcgtggcgcgacg	254
DB	705	CGACGACGACGACGACGAGGAGGACCGCGGTTCGACGACGAGGACGAGGACGAGGA	764
QY	255	gaacgcgaagacaacgg---ccatgccgagggccactgcyacgcgcgcgcgcgcgcgc	311
DB	765	GAACATCGAGAGCAAGGTACCAAGCCCGTGAGATCCAGAAAGCGCGCGTGAAGCGCC	824
QY	312	tgcgaacgcgaagaccaacggccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	371
DB	825	CGCCCCCGGACGAGCAGGACGACGACGAGGAGGACGACGACGACGACGACGACGAC	884

QY 307 gccactgcgaacggcaaacacggccacacgcagagcaacggcgcattgtagggccgc 366
 Db 104640 VCCGNNNNNSGCGCGCCGCGCCGCGGCGNNNNNSGCGCGCGCCGCGCCGCGC 104581
 QY 367 gacgcgaacggcgagcaacagagcatccgagggactccggcgcaacggcgagagcaac 426
 Db 104580 GGNHHNNNSGCGCGCGCCGCGCCGCGGCGNNNNNSGCGCGCGCCGCGCGC 104521
 QY 427 ggg--catgcggcgccgcg--gcaagagagagagggcggtgagtgaaattcgcggt 483
 Db 104520 CGNNHHNNNSGCGCGCGCCGCGCCGCGGCGNNNNNSGCGCGCGCCGCGC 104461
 QY 484 gcaagagcggcggtgctggcgccagcgggcgcaacatgagcatcggcgcatagcgtac 543
 Db 104460 GCCVCGGNNHHNSGCGCGCGCGCGCCGCGCCGCGCGGCGNNNNNSGCGCGCG 104401
 QY 544 aagatcagcgagcgtgtaggagagggcgcgccgctgctgcgtgcgtgcgcacggg 603
 Db 104400 CGCGCCCG 104341
 QY 604 gaccgctcgtttcccgcccttcgcagcggcgccgctcgagggcgagagcgctgcgcgc 663
 Db 104340 GCCCG 104281
 QY 664 gccgtgcgcacggcgcaagttaactgctaccccgcgcgctgcgtcccgccgcacga 723
 Db 104280 GCCVCGGNNHHNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104221
 QY 724 agcgcgtgcgagacactgtcgagggcggtgcgtacatcgtatcgccgacacgctc 783
 Db 104220 CCGGNNHHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104161
 QY 784 ttccctacgcgcggcgacccagcgatcgaggtcataatccgggtgctgcccagacc 843
 Db 104160 GNNHHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104101
 QY 844 gccggcgcaacattctgtctcccgagcgagctaccacaaatacagagcgcgccgcg 903
 Db 104100 GACGCGGNNHHNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104041
 QY 904 ttcaacaggctgaggtc 921
 Db 104040 NNSGCG 104023

RESULT 13

X53491
 ID X53491 standard; DNA; 114955 BP.
 AC X53491;
 DT 05-JUL-1999 (first entry)
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 OS Synthetic.
 PN WO9913886-A1.
 PD 25-MAR-1999.
 PF 17-SEP-1998; U19419.
 PR 09-JUN-1998; US-093972.
 PR 17-SEP-1997; US-059160.
 PA (UYEC-) UNIV EAST CAROLINA.
 PI Nyce JW.
 DR WPI; 99-229400/19.
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 PS Disclosure; Page 37; 120pp; English.

CC The specification describes antisense oligonucleotides (X52869-X53271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences X5272-74. These multiple target
 CC oligonucleotides (specifically X5180-271) can be used for the antisense
 CC treatment of diseases and conditions. Typical diseases and conditions
 CC are those associated with impaired respiration and inflammation,
 CC including lung diseases, pulmonary vasoconstriction, inflammation,
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 CC metastases, as well as all types of cancers which may metastasize or have
 CC metastasized to the lungs, including breast and prostate cancer.
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match 3.6%; Score 69.6; DB 1; Length 114955;

Best Local Similarity 35.0%; Pred. No. 0.00018;

Matches 230; Conservative 44; Mismatches 379; Indels 5; Gaps 1;

QY 95 gcgacggagtcgcgcgaacgcttgcgtggcgacgcgcgaacggcgaagcaacg 154
 Db 108402 CGCGCGCTGCTCGCSNNNDNNGCGCGCGCGCGCTGCTGCSNNNDNNGCGG 108461
 QY 155 gccatggcgtgctgcgcgcgtgaacgcgaagacacgcgcgcgtggtgacgcgcgc 214
 Db 108462 GCGCGCGCGCGCTGCTCSNNNDNNGCGCGCGCGCGCTGCTCSNNNDNNGCG 108521
 QY 215 cgaacggcgaagcaacggccatgctgctgcgcgcgaacgcgcgcgaagcaacgccc 274
 Db 108522 GCGCGCGCGCGCTGCSNNNDNNGCGCGCGCGCGCGCTGCSNNNDNNGCGG 108581
 QY 275 atcgagggccactgcgaacggccacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 334
 Db 108582 GCGCGCGCGCTGCSNNNDNNGCGCGCGCGCGCGCTGCTCSNNNDNNGCGG 108641
 QY 335 accgcgagcaacggccatgctgagccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 394
 Db 108642 CGCSNNNDNNGCGCGCGCGCGCSNNNDNNGCGCGCGCGCGCSNNNDNNGG 108701
 QY 395 cgcgagactccgcgcgaacggcgagagcaacgcgcgcgcgcgcgcgcgcgcgcgcgc 454
 Db 108702 CGGCGCGCGCGCSNNNDNNGCGCGCGCGCGCSNNNDNNGCGCGCGCGCSNN 108761
 QY 455 aggagcggtgagtggaatttcgcggtgcgaagcgcgcgcgcgcgcgcgcgcgcgcgc 514
 Db 108762 NGCGCGCGCGCSNNNDNNGCGCGCGCGCSNNNDNNGCGCGCGCGCGCGCTG 108821
 QY 515 cgaacatgagcatcgc 574
 Db 108822 CGCTTBGGGCCCSNNNDNNGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 108881
 QY 575 cgcgcccgcgtgctgc 629
 Db 108882 NGGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108941
 QY 630 cagcgccgctgagcgcgagggcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 689
 Db 108942 CGCTTBGGGCCCSNNNDNNGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 109001
 QY 690 ctaccccgcggtctggcctcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 747
 Db 109002 CGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 109059

RESULT 14

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 15:03:30 ; Search time 1096.08 Seconds
(without alignments)
3437.285 Million cell updates/sec

Title: us-09-026-400-3
Perfect score: 1910
Sequence: 1 cgcgctactagtagtattcc.....aagttaaaaaaaaaaaaa 1910

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	372.6	19.5	571	49	AI657244	AI657244 486092E01
2	313.8	16.4	594	49	AI619119	AI619119 486073F11
3	256.2	13.4	459	21	D48521	D48521 RICS14756A
4	233.8	12.2	410	20	D25143	D25143 RICS343A R
5	182.4	9.5	464	49	AU056779	AU056779 AU056779
6	151.6	7.9	446	20	Z38084	Z38084 ATTS4224 Ve
7	132	6.9	470	23	T76370	T76370 11148 Lambd
8	121	6.3	352	23	T13684	T13684 1849 Lambda
9	118.4	6.2	343	20	D23043	D23043 RICC2086A R
10	108.8	5.7	283	49	AI621525	AI621525 486092E01
11	102.2	5.4	719	40	AA980369	AA980369 ua52c12.r
12	92.4	4.8	512	34	AA511748	AA511748 v129f11.r
13	89.8	4.7	405	20	Z38047	Z38047 ATTS4207 Ve
14	85.2	4.5	415	48	AI597379	AI597379 v129f11.y
15	80.4	4.2	347	23	T41757	T41757 10338 Lambd
16	78.2	4.1	801	42	AI116859	AI116859 ue29d01.y
17	77	4.0	559	33	AA394723	AA394723 26506 Lam
18	75.6	4.0	729	41	AI048570	AI048570 ud61e01.y
19	75	3.9	757	40	C91185	C91185 C91185 Dicc
20	74.4	3.9	536	47	AI487927	AI487927 EST146249
21	69.8	3.7	664	41	AI055475	AI055475 coau0004B
22	66.6	3.5	870	48	AI562691	AI562691 TENS2678
23	64.6	3.4	779	48	AI562604	AI562604 TENS2761
24	61.8	3.2	521	42	AI116355	AI116355 uf02b01.y
25	61.6	3.2	437	20	T53940	T53940 Yb85b03.r1
26	61.6	3.2	549	26	W43288	W43288 22661 Lambd
27	61.4	3.2	699	45	AI386363	AI386363 mo05a03.y
28	57.6	3.0	528	25	N97159	N97159 22338 Lambd
29	57.2	3.0	820	41	AI001336	AI001336 MEST6-A6
30	56.2	2.9	288	36	AA650709	AA650709 30823 Lam
31	55.4	2.9	470	37	AB009149	AB009149 AB009149
32	54.8	2.9	248	36	C72041	C72041 C72041 Rice
33	54.8	2.9	641	45	AI357868	AI357868 gw13b02.x
34	54.2	2.8	676	45	AI389106	AI389106 GH20192.5
35	53.8	2.8	581	41	AI063350	AI063350 GH03102.3
36	53.4	2.8	806	47	AI523757	AI523757 tg94f08.x
37	53.4	2.8	610	48	AI609782	AI609782 t183a03.x
38	53.4	2.8	598	49	AI658863	AI658863 t198f04.x
39	53	2.8	833	43	AI189624	AI189624 qd32h08.x
40	53	2.8	640	49	AU056191	AU056191 AU056191
41	52.8	2.8	503	43	AI184898	AI184898 oo07h05.x
42	52.8	2.8	539	43	AI192517	AI192517 qe61c01.x
43	52.8	2.8	449	45	AI369183	AI369183 q175b04.x
44	52.8	2.8	599	46	AI453638	AI453638 t145g10.x
45	52.8	2.8	448	50	AI694026	AI694026 wd13h09.x

ALIGNMENTS

RESULT 1

AI657244

LOCUS

DEFINITION

486092E01.y1 486 - leaf primordia cDNA library from Hake lab Zea

Accession

AI657244

NID

94753339

VERSION

AI657244.1 GI:4753339

EST

571 bp

mrna

05-MAY-1999

EST.
Zea mays.
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Zea.
REFERENCE
1. (bases 1 to 571)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
TITLE
Unpublished (1999)
JOURNAL
On Jun 5, 1998 this sequence version replaced gi:3189324.
CONTACT: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 486092 row: E column: 01.
FEATURES
Location/Qualifiers
i. . 571
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="486 - leaf primordia cDNA library from Hake
lab"
/tissue_type="leaf primordia"
/dev_stage="P7-P11 leaf"
/lab_host="E.coli XL1-Blue MFR"
/note="Organ: shoot; Vector: lambda zap; Hake lab cDNA
library."
BASE COUNT
108 a 198 c 178 g 87 t
ORIGIN

QY	1021	cccaacaacccgtcggcagcgtttactactcagcacatctgtccaaggtcgcgagggtg	1080
Db	478	CCCCAACCCCTTGGCGGAGTGTCTACACCCGAGACGATTTGGCCAAGGTTCGCGGA-GTG	536
QY	1081	gcgaagagcgtcggaatatattggtgattgtgcagca	1115
Db	537	GCAAGGAGCTTGGAACTACTGTCATCGCTGATGA	571
RESULT	2		
LOCUS	AI619119/c		
DEFINITION	AI619119 594 bp mRNA	EST	21-APR-1999
	486073f11.x2 486 - leaf primordia cDNA library from Hake lab zea		
ACCESSION	AI619119	mays cDNA, mRNA sequence.	
NID	94528245		
VERSION	AI619119.1	GI:4628245	
KEYWORDS	EST.		
SOURCE	Zeja mays.		
ORGANISM	Zeja mays		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.		
AUTHORS	1 (bases 1 to 594)		
TITLE	Walbot V.		
	Maize ESTs from various cDNA libraries sequenced at Stanford University		
JOURNAL	Unpublished (1999)		
COMMENT	On Mar 10, 1998 this sequence version replaced gi:2948200.		
	Contact: Walbot V		
	Department of Biological Sciences		
	Stanford University		
	855 California Ave, Palo Alto, CA 94304, USA		
	Tel: 650 723 2227		
	Fax: 650 725 8221		
	Email: walbot@stanford.edu		
	Plate: 486073 row: F column: 11.		
FEATURES	Location/Qualifiers		
source	1..594		
	/organism="Zea mays"		
	/cultivar="B73"		
	/db_xref="taxon:4577"		
	/clone_lib="486 - leaf primordia cDNA library from Hake lab"		
	/tissue_type="leaf primordia"		
	/dev_stage="p7-p11 leaf"		
	/lab_host="E.coli XL1-Blue MFR"		
	/note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA library."		
BASE COUNT	172 a	113 c	112 g 197 t
ORIGIN			

Qy	1016	taaaccccaacaaaccgtggcgagctttactoctacgacctatgttccaagtgcgg	1075
Db	61	TAAATCCCAATAATCCCATGCGGGAATGTACACTTACGAGCATTATCCCAAGTGCGAG	120
Qy	1076	aggtagcgaagaagctcggaatatgtgtattgtctgacgaggtatacgcgaagctggcttc	1135
Db	121	AGGTAGCAAGAAGACTTGGGNWTTTGGTAATTACTGATAGGTGTATGGTAATTGGTTT	180
Qy	1136	tggcgagcgccccgttcatcccacaatgggagtggttgggcacatcacccctgtgctgtcca	1195
Db	181	TTGGAGTTCGCCCATTTGTCCCAATGGTTGCTTTGGCACATCGTACCAATATTAAACA	240
Qy	1196	tagggctctgtccaatgatcgttagtgctgtagtggcggttgagtggtgtagcggtgt	1255
Db	241	TAGATCGCATATCAAAGAGGTGGTAGTGC GGATGGCGATTTGGTTGGTAGCAATAT	300
Qy	1256	acgaccacgaagaatcttacagaaaactaagatctctacatcaattacgaattacctcca	1315
Db	301	GTGACCCCAGAAGACTCTACAAAGAACCAAGNTGCAACATTAATTACTTAATTCCTTA	360
Qy	1316	atgtctgacagaccccgacacottcattcaggcagctctctcctcagattcttgagaaca	1375
Db	361	ATGTTTCAACTGTACCAGCACTTTCATTTCAGGAGGCTCTACCGAATATTCTTAAGATAC	420
Qy	1376	caaaggaagattttttaaggcgattattgctgtctaaag	1416
Db	421	CAA--GGGAAGATCTTTAAAGGGGTAATTGATTGCCTTAGC	459
 RESULT 4			
D25143			
LOCUS	D25143	410 bp mRNA	EST 20-JUL-1998
DEFINITION	RICR3343A Rice root Oryza sativa cDNA clone R,	mRNA sequence.	
ACCESSION	D25143		
NID	g428987		
VERSION	D25143.1	GI:428987	
KEYWORDS	EST.		
SOURCE	Oryza sativa.		
ORGANISM	Oryza sativa		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;		
TITLE	Poaceae; Oryza.		
JOURNAL	Minobe,Y. and Sasaki,T.		
COMMENT	Rice cDNA from root Unpublished (1995)		
 Contact: Yuzo Minobe National Institute of Agrobiological Resources Rice Genome Research Program 2-1-2 Kannondai,Tsukuba Ibaraki, Japan 305 Tel: 0298-38-7441 Fax: 0298-38-7468 Email: minobe@rsg.riken.go.jp PROJECT = 'RGP'. Location/Qualifiers 1. .410 /organism="Oryza sativa" /strain="Niponbare, sub_species Japonica" /db_xref="taxon:4530" /clone="R" /clone_lib="Rice root" /note="Prepared from seedling root."			
BASE COUNT	125 a	69 c	97 g 119 t
ORIGIN			
 Query Match 12.2%; Score 233.8; DB 20; Length 410; Best Local Similarity 74.8%; Pred. No. 1.9e-30; Matches 306; Conservative 0; Mismatches 102; Indels 1; Gaps 1;			

Lansing, MI

Plant Physiol. 106, 1241-1255 (1994)

Plant Physiol. 106, 1241-1255 (1994)

MEDLINE 95148729
COMMENT On Nov 29, 1993 this sequence version replaced gi:430053.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: 17.

FEATURES

Location/Qualifiers

1..352

/organism="Arabidopsis thaliana"

/strain="var columbia"

/db_xref="taxon:3702"

/map="4"

/clone="3504T7"

/clone.lib="Lambda-PRL2"

/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."

BASE COUNT 77 a 61 c 83 g 118 t 13 others

ORIGIN

Query Match 6.3%; Score 121; DB 23; Length 352;

Best Local Similarity 63.9%; Pred. No. 1.6e-11;

Matches 175; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 974 tggatccatcgccagacaagaccacccgcgatgtcatcataaaccacacccgt 1033

Db 2 TCGAGCTCTTGACAGAGAAACACCGTTCCTTGGTTGTATAAACCCNGTAACTCTT 61

QY 1034 gcgcagcgtttactctacgaccatctgtccaaagtcgcggagtgccgaaaagctcg 1093

Db 62 GCGGAATNTCTATAGTACCACGATTTGATGAAGATTNCGGAATCGGGAAAACTAG 121

QY 1094 gaatatgtgtatgtacagaggtatagcgaagctggtctctggcagcgcccgcttca 1153

Db 122 GGTTCCTTGTATNCTGATGAGGTTTACGGTCATCTTGTGTTAGCAAAACCGTTTG 181

QY 1154 tcccaatggagtggttgggcacatcacccctgtgtcttcacaggtctcttccaagt 1213

Db 182 TNCAATNGTGTGTTGGTATATGTGTCGTGTTACTCTTGGTCTTATCAAGA 241

QY 1214 catggatgtcctggtgagcggtctggatgggt 1247

Db 242 GATGGATAGTTCAGGTTTCGACTCGGGTTNGT 275

RESULT

9

D23043

LOCUS

D23043 R1C22086A Rice callus Oriza sativa cDNA clone R, mRNA sequence. 20-JUL-1998

DEFINITION D23043

ACCESSION D23043

NID 9426967

VERSION D23043.1 GI:426967

KEYWORDS EST.

SOURCE Oriza sativa.

ORGANISM Oriza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Oriza.

REFERENCE, 1 (bases 1 to 343)

AUTHORS

Sasaki, T. and Minobe, Y.
Rice cDNA from callus
Unpublished (1994)

COMMENT

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

FEATURES

Location/Qualifiers

1..343

/organism="Oriza sativa"

/strain="cultivar Nipponbare, sub_species Japonica"

/db_xref="taxon:4530"

/clone="R"

/clone.lib="Rice callus"

/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."

BASE COUNT 90 a 58 c 80 g 94 t 21 others

ORIGIN

Query Match 6.2%; Score 118.4; DB 20; Length 343;

Best Local Similarity 70.0%; Pred. No. 4.3e-11;

Matches 205; Conservative 0; Mismatches 84; Indels 4; Gaps 4;

QY 1063 tccaaggtcggaggtggtcgaaaaggtcggaattggtgaattggtgagaggtatatac 1122

Db 2 TCCAAGTGGCAGAGTAGCAGGAAGCTTGGGATATGGTAATTAATTACTGATGAGGTGAT 61

QY 1123 gccaagctggttctgtggcagcgcccttcatcccaatggggaggtgtttgggcacatcac 1182

Db 62 GTANTTTGGTTTTGGGAGTTCNCANTTGTCCCAATNGGTACTTTGGGCACATCGTA 121

QY 1183 cctgtgctgcattaggtgtctgtccaaagtcattgtagtagtgc-tgtagtgcgcttg- 1241

Db 122 CCATAANNANCCNTAGATCGCTATCAAGAGGTGGTAGTCCCGTAGATGCGGACTTGA 181

QY 1241 gatgggtagcgtgtacacccccagaaagatcttacagaaactaatctc-tacatca 1299

Db 182 GNTGNTAGCANTAGTGACCCCAAGACACTCTACAAGAAACCAAGATTGCANCCATTA 241

QY 1300 attcgaattacctcaatgtctcgacag-accacgaacaccttcattcaggcgag 1351

Db 242 ATTACTAATTCCTTAATGGTTCAACTGNATCCAGCANCTTTCATTCAGGGAG 294

RESULT

10

AI621525/c

LOCUS

AI621525 283 bp mRNA EST 21-APR-1999
486092E01.xl 486 - leaf primordia cDNA library from Hake lab Zea

DEFINITION

AI621525

ACCESSION

AI621525

NID 94630651

VERSION AI621525.1 GI:4630651

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Zea.

REFERENCE 1 (bases 1 to 283)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

Unpublished (1999)


```

RESULT 12
AA511748      512 bp      mRNA      EST      08-JUL-1997
LOCUS      vj29f11.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
DEFINITION      clone IMAGE:930477 5' similar to gb:X52520_cds1 TYROSINE
AMINOTRANSFERASE (HUMAN);, mRNA sequence.
ACCESSION      AA511748
NID      92249602
VERSION      AA511748.1 GI:2249602
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 512)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1398101.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:535397
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 366.
FEATURES
source
1. .512
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:930477"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dr. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATCGGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTT 3'"
BASE COUNT      118 a      122 c      139 g      133 t
ORIGIN
Query Match      4.8%; Score 92.4; DB 34; Length 512;
Best Local Similarity 52.5%; Pred. No. 1e-06;
Matches 249; Conservative 0; Mismatches 221; Indels 4; Gaps 2;
QY 772 gcgagacgtcttcctaccgcgcggggaccagcgatcgaggtcataatcccggtg 831
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 GCTAAGGATGTCATCTCGAGGAGTGCGTGCAGTCAGGCCCATGTGCTGACGCGTG 63
QY 832 ctgcccagacgcgcgcgcgaacattctgctcccccaggccaggtaccacaaactacgag 891
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 TTGCCC---AATCTGGACAGAACATCTTCATTCGAGGCCCGGGTTTCCCTCTACAGG 120
QY 892 gcgcgcgcgcggttcaacaggtcgaggtccgcgcatcttcacccatcccccagcaagggg 951
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 ACATTGGCTGAGTCTATGGGAGATTGAGGTCAAGCTCTACAATCTATTTCCTGAGAGTCT 180
QY 952 tggagatcacatcgactcgctggaatccatcgccacagaacacacccgcgcgtgc 1011
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 181 TGGAAATTGATCTAAACAACTGGAATCTCTGATCGAGAAAAAACAGCTTGTCTCTG 240
QY 1012 atcataaaaccccaacaccccgctgcgcgagctttactcctacgaccatctcccaaggtc 1071
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GTCAACAAACCCCTCCCAATCCCTGTGGCTGTGTTCAGTAAGCGGCACCTTCAGAAAGAT 300
QY 1072 gcggaggtggcgaaaggctcggaatattggtgattgctgacgaggtatcacgcgaagctg 1131
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TTGGCAGTGGCTGAAGGCAATGCGTCCCATCTTAGCCGATGAGATCTATGGTGACATG 360
QY 1132 gttctgggcagcgccccgttccatcccaatgggagtggtttgggcacatcacccctgtgctg 1191
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTGTTTTCAGATTGCAAAATATGAACCAATGGCCACCCCTCAGCACCATGTCGCCCATCTG 420
QY 1192 tccatagggtctctccaatcatgcatgtagtcctgtagtcgctggtgagtg 1245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TCTGTGGTGGGCTG-GCAAGCGCTGGCTGGTTCCTGGCTGAGGCTGGGCTGG 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
Z38047/c      405 bp      mRNA      EST      12-OCT-1994
LOCUS      ATTS4207 Versailles-VB Arabidopsis thaliana cDNA clone VBMA10 3',
DEFINITION      mRNA sequence.
ACCESSION      Z38047
NID      9557616
VERSION      Z38047.1 GI:557616
KEYWORDS      EST.
SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE      1 (bases 1 to 405)
AUTHORS      CNRS.
TITLE      The Arabidopsis thaliana transcribed genome: the GDR cDNA program
JOURNAL      Unpublished (1996)
COMMENT
Contact: Desprez T., Amselem J., Chiapello H., Rouze P., Caboche
M., Hofte H.
INRA Versailles
Laboratoire de Biologie Cellulaire
Route de Saint-Cyr, 78026 Versailles Cedex, France
Email: thierry@versailles.inra.fr.
FEATURES
Location/Qualifiers
1. 405
/organism="Arabidopsis thaliana"
/strain="ecotype Columbia"
/db_xref="taxon:3702"
/clone="VBMA10"
/clone_lib="Versailles-VB"
/note="Vector: pBluescript; tissue=whole seedlings of A.
thaliana ecotype Columbia; clone library=Versailles-VB;
Cloning vector: pBluescript (Stratagene); Physiological
conditions: in vitro-grown etiolated seedlings, 5 days
old."
BASE COUNT      132 a      80 c      73 g      117 t      3 others
ORIGIN
Query Match      4.7%; Score 89.8; DB 20; Length 405;
Best Local Similarity 62.0%; Pred. No. 2.7e-06;
Matches 142; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 1480 ccagagagatcaatgtttgtcattggtgaactgaactacattctttggaggaaatagac 1539
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 CCAGAAGATCCATGCAATGATGGTTAAGCTGAATCTCTCATTTACTTGAAGATGAAGT 313
QY 1540 gatgacattgattttgtcgaagctcgcaaaagaagaatcagtaaatcttatgtcccaggy 1599
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 GACGATATCGACTCTCTGTTTCAAGTTAGCTAGGGAAGAACTCAGTCATCTCTCTCTGGG 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Qy 1600 agtgtcttggaatggcaaacctgggtccgcattactttgtgtgttccattctctt 1659
Db 252 ACCGGGTGGGCTGAAGAACTGCTGAGGATACGTTTGAGCAGATGCAACTTCGATT 193
Qy 1660 caagatggtctcgaagatcaaatcattctgtcacaagaagaaga 1708
Db 192 GAAGAAGCTTTTAAAGGATCAAAATGTTCTATCTTAGACATGCCAAGA 144

RESULT 14
A1597379 415 bp mRNA EST 21-APR-1999
LOCUS vj29f11.v1 Striatagene mouse diaphragm (#937303) Mus musculus CDNA
DEFINITION clone IMAGE:930477 5', similar to gb:X52520_cds1 TYROSINE
AMINOACIDTRANSFERENCE (HUMAN);, mRNA sequence.
ACCESSION A1597379
NID 94608427
VERSION A1597379.1 GI:4606427
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 415)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189286.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousees@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:535397
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 399.
FEATURES
Source
1..415
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:930477"
/clone_lib="Striatagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally from mRNA
Prepared from diaphragm muscle. Primer: Oligo dt. Average
insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTATTTTATTTT 3'"
BASE COUNT 104 a 102 c 103 g 106 t
ORIGIN
Query Match 4.5%; Score 85.2; DB 48; Length 415;
Best Local Similarity 54.1%; Pred. NO. 1.6e-05;
Matches 196; Conservative 0; Mismatches 163; Indels 3; Gaps 1;
Qy 772 gccgacgacgtcttcctccacgcgcggcgaccgagcgatcgaggtcataatcccggtg 831
Db 4 GCTAAGAGATGTCATCTCTGACGAGTGGCTGTCAGCAGGCCAATGAGCTGTCTAGCCGTG 63

```

```

Qy 832 ctgccccagaccgcgcgcgcacacattctgtctccccagggcaggtaccaccaactacag 891
Db 64 TTGGCC---AATCTGGACAGACATCTCATTCGAGGCCCGGGTTTTCCTCTACAGG 120
Qy 892 gcgcgcgcgcgttcacaggtcgaggtccggcatttcgacctcattcccccgaagaagg 951
Db 121 ACATTGGCTGAGTCTATGGGGATTGAGGTCAAGTCTACAAATCTATTGCTTGAGAAGTCT 180
Qy 952 tgggagatcgacatcgactcgctggaatccatcgcccaagaacaacaccccgccatggtc 1011
Db 181 TGGGAATGTGCTAAACAACACTGGAATCTCTGATCGACGAAAAACAGCTTGTCTCTG 240
Qy 1012 atcataaccccccaaacccgcgcgcgcgttcttactctctacgacccattcttccaaagtc 1071
Db 241 GTCAACAACCCGCTCCAAATCCCTGTGGCTCTGTGTTCAAGTAGCGGCACCTTCAGAAGATT 300
Qy 1072 gcggaggtggcgaagaagctcggaattatgggtgattctgacgaggtatatacggaagctg 1131
Db 301 TTGGCAGTGGCTGAAGGCAATGCGTCCCATCTTAGCCGATGAGATCTATGTTGATGATG 360
Qy 1132 gt 1133
Db 361 GT 362

RESULT 15
T41757
LOCUS T41757 347 bp mRNA EST 07-AUG-1995
DEFINITION 10338 Lambda-PRL2 Arabidopsis thaliana cDNA clone 67A8T7, mRNA
sequence.
ACCESSION T41757
NID 9931391
VERSION T41757.1 GI:931391
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 347)
Newman, T., deRuijn, F.J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomasow, M.,
Retzel, E. and Somerville, C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
On Jan 24, 1995 this sequence version replaced gi:634345.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313t@ibm.cl.msu.edu
Insert Length: 357 Std Error: 0.00
Seq primer: T7 dye primer
High quality sequence stop: 250.
FEATURES
Location/Qualifiers
1..347
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="67A8T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture

```

